**Quiz 2**

*Content: Exploring the association between variables in a data set*

In 1990, 1200 post-menopausal women were recruited for a study to

investigate the effect of a Post-Menopausal Hormone (PMH) on the

incidence of breast cancer (200 users and 1000 non-users).

Given a data set of 1200 women, breast\_cancer.csv, in which we concern the columns below.

bbd: Status having breast cancer, 1= Yes and 0 = No

agemenop (Age): The age of the woman at menopausal

pmh: Status whether or not the woman use PMH (post-menopausal hormone). 2 = No and 3 = Yes

bmi: The body mass index

Purpose: We want to form a model/classifier that help us to predict the status of having breast cancer for women, based on their information of Age, PMH and BMI.

**Q1.** In order to know if we should use variable Age in the model/classifier, we should explore the association between the response variable with Age. True or False?

Ans: TRUE

**Q2.** How to explore the association between the response bbd and age? (MCQ)

Box plot of age for two groups of bbd

Histogram of age

Correlation coefficient of bbd and age

Scatter plot for bbd and age

ANS: Box plot of age for two groups of bbd

**Q3.** How to explore the association between the response bbd and pmh (multiple answers)

Box plot of pmh for two groups of bbd

Contingency table with conditional probabilities of having breast cancer for different categories of pmh

Odds ratio

Correlation coefficient of pmh and bbd

Ans: Contingency table and Odds ratio.

**Q4.** Given the photo. Which of the following statements is correct.

The probability of not having cancer (bbd = 0) is 1.4 times the probability of having cancer.

The probability of not having cancer (bbd = 0) among non-PMH users (pmh = 2) is 1.4 times the probability of not having cancer among PMH users (pmh = 3).

The probability of having cancer (bbd = 0) among non-PMH users (pmh = 2) is 1.4 times the probability of having cancer among PMH users (pmh = 3).

The odds of not having cancer (bbd = 0) among non-PMH users (pmh = 2) is 1.4 times the odds of not having cancer among PMH users (pmh = 3).

ANS: The odds of not having cancer (bbd = 0) among non-PMH users (pmh = 2) is 1.4 times the odds of not having cancer among PMH users (pmh = 3).

A screenshot of a computer

Description automatically generated